

研究收入分配问题的一种 新洛伦兹曲线模型:建构与应用

游士兵 王原君*

摘要:洛伦兹曲线模型研究是近年来收入分配领域较为活跃的研究领域之一,本文先提出了几种应用效果良好的模型,用国外文献中通常采用的方法与国际上通用的数据对这些模型的应用效果进行了分析,说明新模型远优于利用现有文献中模型构造的模型。然后将其中效果最好的模型用于研究我国的收入分配。本文使用了湖北省2009年、2010年、2012年三年的收入分配抽样数据,用其形成分组数据,然后用新洛伦兹曲线模型对产生的分组数据进行拟合,再利用拟合得到的模型估算基尼系数,证明由此得到的基尼系数与从样本数据计算得到的基尼系数基本相同。估算得到的洛伦兹曲线几乎准确地通过分组数据点。说明本文提出的洛伦兹曲线模型具有重要的应用价值,适用于研究我国的收入分配。对湖北省收入分配数据的分析说明,我国城乡收入分配中财产性收入不平等最严重,甚至部分人口尚无财产性收入。农村转移性收入的数量或比例还落后于城镇。

关键词:洛伦兹曲线 收入分配 基尼系数

一、引言

洛伦兹曲线模型是收入分配理论与应用中极为重要的模型,主要用于从分组数据出发构造收入分配的统计分布。国际上部分发达国家已经形成了收入分配抽样数据库,因此可以使用例如Kernel法产生收入分配统计分布。Kernel法的优势是大样本收敛性,即当样本量趋于无穷时,相应统计分布依概率收敛于相应总体的统计分布。但对很多国家,可用的收入分配数据仅是所谓分组数据,这种数据中仅给出了洛伦兹曲线上的若干个(通常10个点左右)。经济学界已经展开了多年研究,企图寻找合适的模型逼近这种分组数据,从中得到相应的统计分布,进而可以使用经济理论中已经存在的理论与方法进行收入分配分析。国内有关洛伦兹曲线模型的研究还处于起步阶段,本文的目的:一是向国内收入分配理论与应用工作者介绍分组数据方法,二是提出几种应用效果良好的分组数据模型,三是用数据试验说明这些新模型适用于我国的收入分配分析。

*游士兵,武汉大学经济与管理学院,邮政编码:430072,电子信箱 sbyou@whu.edu.cn;王原君,武汉大学经济与管理学院,邮政编码:430072,电子信箱:277164479@qq.com。

本文研究得到国家社科基金重点项目“收入分配体制改革对改善居民收入的效果评估研究”(项目号:12AZD030)、教育部规划基金项目“中国收入分配两级分化测算的理论、模型与应用研究”(项目号:09YJA790152)的支持。非常感谢匿名审稿人提出的宝贵修改意见,感谢编辑部的老师们对本文所做的工作。

设 $l(p)$ 是收入分配 X 的洛伦兹曲线, 相应收入分配密度与分布函数分别为 $f(x)$ 与 $F(x)$, μ 是平均收入, 则按定义有:

$$l(p) = \frac{1}{\mu} \int_0^x tf(t) dt, p = F(x) \quad (1)$$

可见应有:

$$l'(p) = \frac{x}{\mu} \quad (2)$$

$$f(x) = \frac{1}{\mu l''(p)} \quad (3)$$

注意到 p 与 x 满足关系式 $p = F(x)$ 。

但实践中收入分配 X 是未知的, 可以通过抽样调查对未知总体进行分析。但各国(包括我国)通常只发布分组数据, 完整分组数据形如:

$$(p_i, L_i), i = 1, 2, \dots, n \quad (4)$$

$$(p_i, x_i/\mu), i = 1, 2, \dots, n \quad (5)$$

其中 p_i 是收入低于 x_i 的人口比例, L_i 是这一低收入群体拥有的总收入比例, μ 是平均收入。很多国家只发布(4)型的或只发布(5)型的数据。已知数据背后的洛伦兹曲线为 $l(p)$ 时, 上述数据点必满足:

$$l(p_i) = L_i$$

$$l'(p_i) = \frac{x_i}{\mu}$$

分组数据的形成分成两步, 一是将抽样调查得到的收入数据从低到高排序, 然后分组, 取若干最低收入者形成第一组, 之后在剩下的人口中取若干最低收入者形成第二组, 等等, 形成 n 组, 最后剩下的人口形成第 $n+1$ 组。第二步是作统计计算, 取第 i 组中最低收入数量作为 x_{i-1} , 记 $x_0 = 0$, 记所有人口的平均收入为 μ , 即可形成(4)式与(5)式。注意到, 即使是抽样数据能够得到, 仍无法准确得到总体的统计分布, 除非是把所有人口的收入信息都收集起来。因此, 即使使用例如 Kernel 法从抽样数据出发估计统计分布, 得到的仍只是总体分布的有限样本近似。

由于实践中只有分组数据可用, 经济理论界通过构造模型 $L(p, \tau)$ 来估计 $l(p)$, 这种模型即洛伦兹曲线模型。 $L(p, \tau)$ 是含有参数向量 τ 的函数, τ 的作用是调节曲线的形状以更好地模拟给定的数据。若构造了模型 $L(p, \tau)$, 通常用最小二乘法拟合数据点列(4), 即通过极小化:

$$\sum_{i=1}^n [L_i - L(p_i, \tau)]^2 \quad (6)$$

来确定参数向量 τ 的估计值 $\hat{\tau}$, 从而得到 $l(p)$ 的估计 $L(p) = L(p, \hat{\tau})$ 。为简单, 下面略去 τ 或 $\hat{\tau}$ 使表述简单。显然模型 $L(p)$ 应满足条件:

$$L(0) = 0, L(1) = 1, L'(p) \geq 0, L''(p) \geq 0 \quad (7)$$

用 $L(p)$ 代替(2)式与(3)式中的 $l(p)$ 来求得 $f(x)$ 的估计, 仍记为 $f(x)$, 可见应先求出方程:

$$\mu L'(p) - x = 0 \quad (8)$$

的根 p , 即 $F(x)$ 的近似值, 将 p 值代入 $L''(p)$, 然后再由(3)式得到:

$$f(x) = \frac{1}{\mu L''(p)} \quad (9)$$

以此作为收入分配密度的近似值。

可见 $L''(p) \geq 0$ 这一条件是重要的,这时对于任何固定的 x , (8) 式的左端是增函数,从而保证从中能够解出 p , 而(9)式也保证 $f(x)$ 是正的。于是,只要得到了洛伦兹曲线的估计,就容易算得收入分配的统计分布。Kakwani (1980) 指出,利用分组数据进行收入分配分析是可行的,而且精度是可以接受的,例如,与利用抽样数据估算的基尼系数比较,用分组数据算得的基尼系数的误差在 1% 以下。Wang 等(2011)也说明,与利用抽样数据及 Kernel 法得到的统计分布比较,仅用分组数据算得的统计分布在某些局部对原始数据的逼近更有优势。这是毫不奇怪的,因为 Kernel 法仅具有大样本收敛性,而这种收敛性并不能保证有限样本条件下对原始数据的良好逼近。因此,在很多理论分析与实际应用中,分组数据已经足够且可能比利用样本数据的 Kernel 法更好。因此,洛伦兹曲线模型的研究具有重要意义。

二、洛伦兹曲线模型研究进展

有关洛伦兹曲线模型研究的较早文献是 Kakwani 和 Podder (1973), 经过至今 40 多年的努力,直到 2010 年,经济理论界找到的模型并不多,且大多只含有 1-3 个参数,应用效果一般不太理想。文献作者通常提出一种模型,然后利用分组数据对模型的精度进行考察。最著名的洛伦兹曲线模型是:

$$L(p) = 1 - (1-p)^\beta, \beta \in (0, 1] \quad (10)$$

它是帕累托分布对应的洛伦兹曲线, Kakwani (1980) 给出了该模型的详尽推导。Kakwani 和 Podder (1973) 提出的模型为:

$$L(p) = pe^{-\gamma(1-p)}, \gamma > 0 \quad (11)$$

与:

$$L(p) = p^\alpha e^{-\gamma(1-p)}, 1 \leq \alpha < 2 \quad (12)$$

Basmann 等(1990)将(12)式推广成含四个参数的模型:

$$L(p) = p^{ap+b} e^{-g(1-p^2) - h(1-p)} \quad (13)$$

并利用这一模型推出了研究美国收入分配的专著,见 Basmann 等(1991)。Kakwani (1980) 提出了一种复杂的模型,先考虑将洛伦兹曲线所在的 (p, L) 坐标系转换到 (π, η) 坐标系:

$$\pi = \frac{1}{\sqrt{2}}(p+L), \eta = \frac{1}{\sqrt{2}}(p-L)$$

以避免洛伦兹曲线在 $p=1$ 处的奇异性,对分组数据进行类似转换后,使用模型:

$$\eta = a\pi^\alpha (\sqrt{2} - \pi)^\beta \quad (14)$$

对其进行拟合,这是一种三个参数的模型。Kakwani (1986) 提出了另一种三参数模型:

$$L(p) = p - ap^\alpha (1-p)^\beta, 0 < \alpha \leq 1, 0 < \beta \leq 1, a \text{ 不定} \quad (15)$$

Schader 和 Schmid (1994) 建议将此模型改进为:

$$L(p) = p^\gamma - ap^\alpha (1-p)^\beta \quad (16)$$

实际应用发现(14)-(16)式的精度非常好,这些发现以及实践的需要推动了理论工作者寻找更合理模型的研究。但(13)-(16)式的重要缺陷是它们不满足洛伦兹曲线的条件,因此估计得到的洛伦兹曲线可能不具有合理的经济意义。Ryu 和 Slottje (1996) 建议用伯恩斯坦多项式

逼近洛伦兹曲线,但只能得到所谓“随机凸”的模型,因此也不严格满足洛伦兹曲线的条件。

模型(10) - (12)满足洛伦兹曲线的条件,但精度通常很差。其他一些学者提出了若干种满足洛伦兹曲线条件的模型,Chotikapanich (1993)提出的模型为:

$$L(p) = \frac{e^{\lambda p} - 1}{e^{\lambda} - 1}, \lambda > 0 \quad (17)$$

Gupta (1984)提出了模型:

$$L(p) = pA^{p-1}$$

Rohde (2009)提出了模型:

$$L(p) = \frac{(\phi - 1)p}{1 - \phi p}$$

可见这三种模型都满足洛伦兹曲线的条件,但由于都只有一个参数,应用效果自然不可能很好。Ortega 等(1991)提出了含两个参数的模型:

$$L(p) = p^{\alpha}(1 - (1 - p)^{\beta}), \alpha \geq 0, \beta \in (0, 1]$$

Rasche 等(1980)、Ogwang 和 Rao (1996)、Ogwang 和 Rao (2000)、Villasenor 和 Arnold (1989)、Sarabia(1997)、Sarabia 等(1999, 2001, 2005)、Wang 等(2011)等相应提出了各自的模型。其中 Sarabia 等(1999)将上式推广成:

$$L(p) = p^{\alpha}(1 - (1 - p)^{\beta})^{\eta}, \alpha \geq 0, \beta \in (0, 1], \eta \geq 1$$

张奎和王原君(2010)将其进一步推广成:

$$L(p) = p^{\alpha}(1 - (1 - p^{\gamma})^{\beta})^{\eta}$$

上述这些模型都满足洛伦兹曲线的条件。其中 Wang 等(2011)的数值试验说明,利用有关模型估计收入分配统计分布时,估算结果甚至优于依赖抽样数据的 Kernel 法,这对于收入分配分析具有重要意义,因为研究大多数国家的收入分配时缺少的正是抽样数据。

洛伦兹曲线模型在各国的收入分配中得到了广泛应用,例如世界银行的经济学家为了研究全球收入不平等与贫困问题,建立了软件系统 POVCAL,其中使用了模型(15)与 Villasenor 和 Arnold(1989)提出的一种三参数模型。该软件被大量用于各国的贫困问题分析。

以上这些国外文献都出现于国际上重要的经济理论杂志。在 Wang 等(2011)出现之前,经济理论文献中总共出现了 20 种左右的模型,Schader 和 Schmid (1993)对其中一些模型进行了比较研究。由于经济制度、风俗习惯、人文背景等诸多差别,世界各国收入分配统计分布是极为多样的,任何新的模型都具有重要意义。

分组数据模型对我国具有特别重要的意义。我国的收入分配数据城乡分列,统计部门仅在《中国统计年鉴》中发布分组数据,理论与应用分析人员通常不能得到抽样数据。因此,研究我国的收入分配时,精度良好的分组数据模型不可或缺。同时,我国其他很多部门编撰出版的统计数据中存在大量分组数据,利用分组数据模型可以有效地分析这些数据,从而对我国经济的很多方面展开深入分析。实际上,洛伦兹曲线模型研究不仅对收入分配问题具有重要的理论与应用价值,对工程技术问题中需要估计统计分布的问题也可能使用洛伦兹曲线法。

国内经济理论文献中有关洛伦兹曲线模型的研究极少。本文先提出三种新的单参数模型,如上述文献可见,这三种模型本身具有重要意义,使实际应用中有更多可选的模型。文中再综合利用 Ogwang 和 Rao (2000)提出的杂合(hybrid)技巧与 Wang 等(2011)所提出的一般方法,构造一种洛伦兹曲线模型,并以实际数据说明模型的精度,最后用这种模型对我国的收入分配数据进行实际应用分析。

三、若干新的洛伦兹曲线模型

Ogwang 和 Rao (2000)提出了杂合型方法即凸组合法与乘法加权法,例如考虑(10)式、(11)式、(17)式三者的凸组合:

$$L_o(p) = \delta_1(1 - (1 - p)^\beta) + \delta_2 \frac{e^{\lambda p} - 1}{e^\lambda - 1} + (1 - \delta_1 - \delta_2)pe^{-\gamma(1-p)}$$

$$\beta \in (0, 1], \lambda > 0, \gamma > 0, \delta_1, \delta_2 \in [0, 1], \delta_1 + \delta_2 \leq 1$$

Wang 和 Russell (2013)指出构造组合模型时需要有所选择,例如上述五个参数的模型未必效果良好,尽管其中较多的参数增加了其灵活性。Wang 和 Russell (2013)先构造了一种双参数有理函数模型,之后研究了该模型的曲率(即二次导数与一次导数的比)结构,选择曲率递减的子模型 $L_i(p)$ 与曲率递增的子模型 $L_j(p)$ 构成凸组合,认为这种模型效果应比较理想。一般模型是:

$$\delta L_i(p) + (1 - \delta)L_j(p)$$

这种模型效果良好的原因是用曲率递减的模型拟合数据时,在 $p=0$ 附近会出现下偏,在 $p=1$ 附近会出现上偏;用曲率递增的模型拟合数据时,会出现恰好相反的偏,两种模型中和的结果恰好能达到理想的拟合效果。该文指出,再选择一种曲率递增的模型 $L_m(p)$ 构造乘法加权模型:

$$L_m(p)^\alpha [\delta L_i(p) + (1 - \delta)L_j(p)]^\beta$$

发现这种模型效果极好,用其对世界各国的数据进行拟合时,都能够产生极好的逼近。其中关键点是 $L_i''(p) \geq 0, L_j''(p) \geq 0, L_m''(p)/L_m'(p)$ 递增,使上述模型的参数满足:

$$\alpha \geq 0, \beta \geq 1/2, \alpha + \beta \geq 1$$

这时模型的参数范围达到了某种最大化,模型更为灵活,从而达到理想的应用效果。本文说明,构造上述类型的模型时,可以直接根据 $L_i(p)$ 与 $L_j(p)$ 的导数来了解它们可能出现相对的偏,这使构造类似模型更简单。

实践中的洛伦兹曲线 $l(p)$ 常常是 $l'(0)$ 极小,但不一定有 $l'(0) = 0$ 。我们提出的方法是,寻找满足 $L_i'(0) = 0$ 的模型 $L_i(p)$,将 $L_i(p)$ 用于相应分组数据时,可能在 $p=0$ 处出现下偏。但 $L_i(p)$ 在 $p=0$ 处下偏意味着 $L_i(p)$ 在 $p=0$ 处凸向 p 轴,从而位置过低,而 $L_i(1) = 1$ 的限制需要它很快上升,这样就可能使其在 $p=1$ 处出现上偏。再找模型 $L_j(p)$ 使其 $L_j'(0) > 0$,它可能满足 $L_j'(0) > l'(0)$,于是可能出现与 $L_i(p)$ 相反的偏。这样, $\delta L_i(p) + (1 - \delta)L_j(p)$ 就可能是非常理想的模型。当然,恰好满足这种条件的 $L_i(p)$ 与 $L_j(p)$ 可能存在,但构造这种模型是值得研究的课题。我们从简单模型出发来解决这一问题。

先构造三种新的对数函数模型,它们都是单参数模型。

模型一,考虑:

$$L_\omega(p) = \left(\ln \frac{1 + \omega}{1 - \omega} \right)^{-1} p \ln \frac{1 + \omega p}{1 - \omega p} = \phi p \ln \frac{1 + \omega p}{1 - \omega p} \quad (18)$$

其中 $\phi = \left(\ln \frac{1 + \omega}{1 - \omega} \right)^{-1}$, ω 是其中唯一的参数。

引理 1: 当 $0 \leq \omega \leq 1$ 时 $L_\omega(p)$ 满足洛伦兹曲线的条件,且 $L_\omega'''(p) \geq 0$ 。

证明:先注意到 ϕ 恒为正数。显然 $L(0) = 0$ 与 $L(1) = 1$ 成立。由于:

$$L'_\omega(p) = \phi \ln \frac{1 + \omega p}{1 - \omega p} + 2\omega\phi p \frac{1}{1 - \omega^2 p^2}$$

可见 $0 \leq \omega < 1$ 时, 对任何 $p \in [0, 1]$ 有 $L'_\omega \geq 0$ 。又由于:

$$L''_\omega(p) = 4\omega\phi \frac{1}{1 - \omega^2 p^2} + 4\phi\omega^3 \frac{p^2}{(1 - \omega^2 p^2)^2}$$

可见 $L''_\omega(p) \geq 0$ 对任何 $p \in [0, 1]$ 成立。因此证明 $L_\omega(p)$ 是洛伦兹曲线。而显然有:

$$L'''_\omega(p) = 4\omega\phi \frac{2\omega^2 p}{(1 - \omega^2 p^2)^2} + 4\phi\omega^3 \frac{2p(1 - \omega^2 p^2) + 4\omega^2 p^3}{(1 - \omega^2 p^2)^3} \geq 0$$

证毕。

模型二, 考虑:

$$L_\tau(p) = \left(\ln \frac{1 + \tau}{1 - \tau} \right)^{-1} \ln \frac{1 + \tau p}{1 - \tau p} = \phi \ln \frac{1 + \tau p}{1 - \tau p} \quad (19)$$

其中 $\tau \in (0, 1)$, $\phi = \left(\ln \frac{1 + \tau}{1 - \tau} \right)^{-1}$ 。可见 $L_\tau(0) = 0, L_\tau(1) = 1$ 成立, 且:

$$L'_\tau(p) = \phi \frac{1 - \tau p}{1 + \tau p} \cdot \frac{\tau(1 - \tau p) + (1 + \tau p)\tau}{(1 - \tau p)^2} = \phi \frac{2\tau}{1 - \tau^2 p^2} \geq 0$$

$$L''_\tau(p) = \phi \frac{4\tau^3 p}{(1 - \tau^2 p^2)^2} = \frac{2\tau^2 p}{1 - \tau^2 p^2} L'_\tau(p) \geq 0$$

显然有:

$$L'''_\tau(p) \geq 0$$

从而得到:

引理 2: 当 $\tau \in (0, 1)$ 时 $L_\tau(p)$ 满足洛伦兹曲线的条件, 且 $L'''_\tau(p) \geq 0$ 。

模型三, 考虑:

$$L_\gamma(p) = \frac{\ln(1 - \gamma p)}{\ln(1 - \gamma)} \quad (20)$$

这时:

$$L'_\gamma(p) = -\frac{1}{\ln(1 - \gamma)} \cdot \frac{\gamma}{1 - \gamma p}$$

$$L''_\gamma(p) = \frac{-1}{\ln(1 - \gamma)} \cdot \frac{\gamma^2}{(1 - \gamma p)^2}$$

可见当 $\gamma \in (0, 1)$ 时 $L_\gamma(p)$ 是洛伦兹曲线。显然 $L'''_\gamma(p) \geq 0$, 且:

$$\frac{L''_\gamma}{L'_\gamma} = \frac{\gamma}{1 - \gamma p}$$

显然是 p 的增函数。

引理 3: 当 $\gamma \in (0, 1)$ 时 $L_\gamma(p)$ 满足洛伦兹曲线的条件, 且 L''_γ/L'_γ 对任何 $p \in [0, 1]$ 是增函数。

注意到, 模型一与模型二有本质差别, 有 $L'_\omega(0) = 0, L'_\tau(0) = 2\tau\phi > 0$, 因此如上述使用 L_ω 时在 $p = 0$ 附近容易出现下偏, 由于 $L'''_\omega(p) > 0$, 于是 $L'_\omega(p)$ 以递增的速度增加, 从而 $L_\omega(p)$ 将以极快的速度增加到 $L(1) = 1$, 这意味着在 $p = 1$ 处可能出现上偏; 由于实际应用中的洛伦兹曲线在 $p = 0$ 处的导数值不会很大, 一般会出现 $L'_\tau(0)$ 大于实际洛伦兹曲线的导数值, 于是使用 L_τ 时在 $p = 0$ 附近容易出现上偏, 而 $L_\tau(p)$ 也必须上升到 $L_\tau(1) = 1$, 于是在 $p = 1$ 处就可能出现下偏, 两模型的凸组合互相补偿, 有可能得到良好逼近效果的模型:

$$L_1(p) = \delta L_\omega(p) + (1 - \delta) L_\tau(p) \quad (21)$$

根据 Wang 和 Russell (2013) 的阐述, 可以预料:

$$L_{II}(p) = L_{\gamma}(p)^{\alpha} L_I(p)^{\eta} \quad (22)$$

的应用效果会更好。由引理 1-3, 知对 L_{γ} 来说, 相应 L''/L' 是增函数, 而显然 $L'''_I(p) \geq 0$, 按照 Wang 等(2011)的一般方法, 有:

定理一, 当:

$$\alpha \geq 0, \eta \geq 1/2, \alpha + \eta \geq 1, \delta \in [0, 1] \quad (23)$$

时, $L_{II}(p)$ 满足洛伦兹曲线的条件。

至此, 连同所构造的三种简单模型, 我们共提出了五种新的模型。(21) 式含有三个参数, (22) 含有 6 个参数。下面将说明(21)式的应用效果通常非常好, 其重要缺陷是利用它形成统计分布时对收入 x 有限制, 因为利用(9)式估计密度函数 $f(x)$ 时, 需要先求解 $\mu L'_I(p) - x = 0$ 得到估计值 $p = \hat{F}(x)$, 但 $L'_I(0) > 0$, 且 $L'_I(p)$ 是增函数, 从而对比较小的 x , 可能出现对任何 $p \in [0, 1]$ 恒有 $\mu L'_I(p) - x > 0$ 的情况, 这时对满足 $x < \mu L'_I(0)$ 的 x 就不能估计统计密度。因此, 需要估计统计分布时应该使用模型(22)。另外, 如果仅仅需要估计例如基尼系数, 而不需要估计密度, 使用(21)式将是非常好的选择。

注意到:

$$\lim_{\gamma \rightarrow 0} L_{\gamma}(p) = p$$

因此还可以得到两种相关的模型, 一是(22)式的极限情形:

$$L_{III}(p) = p^{\alpha} [\delta L_{\omega}(p) + (1 - \delta) L_{\tau}(p)]^{\eta} \quad (24)$$

这是一种含五个参数的模型, 其中参数仍应满足条件(23)。若取定 $\eta = 1$, 则得到四个参数的模型:

$$L_{IV}(p) = p^{\alpha} [\delta L_{\omega}(p) + (1 - \delta) L_{\tau}(p)] \quad (25)$$

其中 $\alpha \geq 0$ 。此模型与 Ortega 等(1991)的模型类似, 但它含有 4 个参数, 因此比 Ortega 等(1991)的模型更为灵活。与(21)式比较, 这一模型的优势是对任何 $x > 0$ 都可以计算统计密度。至此, 我们共提出了 7 种新的洛伦兹曲线模型, 扩大了理论与实际应用中洛伦兹曲线模型的选择范围。而且说明了利用其中一些模型, 仅需要信息非常少、稀疏的分组数据, 就得到数据背后收入分配的统计分布。因此, 我们实际上也提出了新的收入分配统计分布的估算方法。

四、模型精度比较

有关洛伦兹曲线模型的研究文献中, 一般使用一些常用的数据进行模型的比较, 这样做的理由是针对相同的数据才有比较的基础。本文使用 Chotikapanich 等(2007)中的数据对所提出的模型进行精度检验, Rohde (2009)使用这些数据对其提出的单参数模型进行了数值试验。本文将只比较 $L_o(p)$ 、 $L_I(p)$ 、 $L_{II}(p)$ 、 $L_{III}(p)$ 、 $L_{IV}(p)$ 。Chotikapanich 等(2007)中列出了涉及 8 个东亚国家或地区 1988 年与 1993 年两年的收入分配分组数据, 每个国家或地区的数据中, 列出了从低到高各个 10% 人口组拥有的总收入比例, 由此可以算得各个国家或地区的洛伦兹曲线上的点列。本文使用 1993 年的中国香港、马来西亚、新加坡三组数据, 拟合结果见表 1-3。我们用:

$$MSE = \frac{1}{n} \sum (L(p_i) - L_i)^2, MAS = \max_i |L(p_i) - L_i|$$

对模型进行比较, 前者是均方差, 后者是最大绝对误差, 可见两者都从全局的角度反映估计值与真实值的差异。这两种误差指标也是国际上研究洛伦兹曲线模型时常用的, L_i 是给定的洛伦兹曲线上的纵坐标。

表 1 对中国香港数据的拟合

p_i	L_i	$L_I(p)$	$L_{II}(p)$	$L_{III}(p)$	$L_{IV}(p)$	$L_o(p)$
0.10	0.0160	0.0188	0.0162	0.0162	0.0162	0.0210
0.20	0.0440	0.0453	0.0437	0.0437	0.0437	0.0478
0.30	0.0800	0.0802	0.0800	0.0800	0.0801	0.0818
0.40	0.1250	0.1242	0.1252	0.1252	0.1252	0.1244
0.50	0.1800	0.1786	0.1799	0.1799	0.1800	0.1775
0.60	0.2460	0.2450	0.2459	0.2459	0.2459	0.2436
0.70	0.3260	0.3266	0.3261	0.3261	0.3260	0.3260
0.80	0.4270	0.4289	0.4270	0.4270	0.4270	0.4302
0.90	0.5680	0.5670	0.5680	0.5680	0.5680	0.5671
1.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MSE $\times 10^6$		2.0107	0.0189	0.0190	0.0215	7.4394
MAS		0.0028	0.0003	0.0003	0.0003	0.0050
Gini Index		0.5186	0.5167	0.5167	0.5169	0.5234

表 2 对马来西亚数据的拟合

p_i	L_i	$L_I(p)$	$L_{II}(p)$	$L_{III}(p)$	$L_{IV}(p)$	$L_o(p)$
0.10	0.0184	0.0203	0.0183	0.0185	0.0180	0.0223
0.20	0.0477	0.0488	0.0478	0.0478	0.0476	0.0510
0.30	0.0863	0.0861	0.0862	0.0861	0.0861	0.0874
0.40	0.1340	0.1332	0.1340	0.1339	0.1341	0.1333
0.50	0.1924	0.1916	0.1924	0.1925	0.1927	0.1907
0.60	0.2639	0.2637	0.2640	0.2641	0.2642	0.2624
0.70	0.3527	0.3531	0.3526	0.3527	0.3526	0.3526
0.80	0.4662	0.4671	0.4663	0.4660	0.4658	0.4682
0.90	0.6240	0.6235	0.6240	0.6240	0.6242	0.6233
1.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MSE $\times 10^6$		0.8403	0.0025	0.0188	0.0624	4.2354
MAS		0.0008	0.0000	0.0001	0.0002	0.0017
Gini Index		0.4768	0.4770	0.4764	0.4763	0.4809

表 3 对新加坡数据的拟合

p_i	L_i	$L_I(p)$	$L_{II}(p)$	$L_{III}(p)$	$L_{IV}(p)$	$L_o(p)$
0.10	0.0216	0.0240	0.0216	0.0218	0.0214	0.0262
0.20	0.0558	0.0569	0.0558	0.0558	0.0556	0.0593
0.30	0.0996	0.0993	0.0996	0.0995	0.0996	0.1007
0.40	0.1535	0.1523	0.1533	0.1533	0.1535	0.1522
0.50	0.2178	0.2171	0.2181	0.2181	0.2183	0.2158
0.60	0.2962	0.2958	0.2961	0.2962	0.2962	0.2942
0.70	0.3912	0.3917	0.3911	0.3911	0.3909	0.3911
0.80	0.5100	0.5111	0.5101	0.5100	0.5098	0.5126
0.90	0.6710	0.6704	0.6710	0.6711	0.6712	0.6705
1.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MSE $\times 10^6$		1.1828	0.0130	0.0207	0.0517	5.5784
MAS		0.0023	0.0003	0.0003	0.0005	0.0045
Gini Index		0.4267	0.4268	0.4267	0.4267	0.4302

上述表1-3中第一、二列是原始数据给出的洛伦兹曲线上的数据点 (p_i, L_i) , 随后五列是用各个模型估算得到的洛伦兹曲线值。表中倒数第三行给出了MSE值, 由于这些值太小, 都将其乘了 10^6 。倒数第二行是MAS, 最后一行是用模型估计的基尼系数。将各模型算得的估计值与第二列的原始数据比较, 可见对中国香港的拟合中, L_{II} 、 L_{III} 、 L_{IV} 三者相差不大。实际上, 这三个模型用于上述三组数据时, MAS都是小数点后第四位的数量级, 这意味着对原始数据拟合时, 误差要低于0.1%。对其他两个国家的数据拟合中, $L_{II}(p)$ 的效果要优于其他各模型。五个参数的模型 L_o 总是不及其他模型, 它甚至不及只含三个参数的 L_I 。因此说明本文的杂合型方法构造的模型是成功的。由于 $L_{II}(p)$ 是最好的, 下面仅使用这一模型分析湖北省的数据。

五、湖北省收入数据分析

湖北省属于中部地区, 经济发展在全国处于中等水平, 分析该省数据在全国有一定的代表性。另外, 我国正在实施到2020年居民收入翻番计划, 从这一点说, 哪怕是对一个省的数据进行较为细致的分析也具有重要意义。由于我们的结果是建立在与抽样数据比较的基础之上的, 所以下面的分析结果具有重要意义。

我们使用湖北省2009年、2010年、2012年三年的收入分组数据进行考察, 数据由湖北省调查总队提供, 含家庭人口等信息。收入数据中给出总收入及工资性收入、经营性收入、财产性收入、转移性收入等四个子项。

对城镇的数据, 按总收入排序后分成10组, 每组占总人口的10%, 最高收入组再等分为两组, 从而总共形成洛伦兹曲线上的10个内部点, 即得到表4中最左边两列的分组数据 $\{(p_i, L_i)\}_{i=1}^{10}$ 。用洛伦兹曲线模型 $L_{II}(p)$ 对此分组数据进行拟合, 然后算得 $\hat{L}_i = L_{II}(p_i)$, 即得以 \hat{L}_i 为表头的第三列。同理, 按工资性收入对人口进行同样分组, 又得到分组数据, 相应洛伦兹曲线上的 L_i 值即表4的第四列; 用 $L_{II}(p)$ 对该分组数据拟合, 又可算得拟合值即表4的第五列 \hat{L}_i 。对经营性收入、财产性收入、转移性收入等都进行类似处理, 得到表4第七、九、十一列洛伦兹曲线估计值。可见依MSE与MAS两种误差测度标准, 各列估计值与分组数据点对应的值都非常接近。

表4的倒数第6行给出了各种收入的平均值, 这一数值是利用样本数据算得的。倒数第三行与第二行分别给出了由模型估计的基尼系数与由样本算得的基尼系数, 可见两种方法算得的基尼系数几乎没有差别。其中对经营收入的洛伦兹曲线估计中误差较大, 但对基尼系数的估计与样本基尼系数相差也是极小的。可见总收入的基尼系数不大, 只有0.256左右。工资性收入的基尼系数稍大一点, 达到了0.324。转移性收入的基尼系数不算小, 达到了0.478。经营性收入的基尼系数很大, 达到了0.75左右, 但这与职业分布有关, 例如国家公务员可能完全没有经营性收入, 因此这一收入的基尼系数很大不能说明什么问题。财产性收入的基尼系数达到极大的0.85左右, 观察洛伦兹曲线可见只有40%左右的人口拥有财产性收入, 其他60%人口的财产性收入为零, 导致基尼系数极大。实际上, 财产性收入在城镇人口收入中的占比还很小, 平均每人不足500元。观察平均收入可见, 城镇居民收入中占大头的是工资性收入与转移性收入, 因此, 这两类收入对总收入的不平等程度具有决定性影响。

经济理论界考虑了如何评价各收入单项(factor component)对总收入不平等的贡献。依

Shorrocks(1982)的分析,在一定条件下,第 k 种单项收入对总收入的不平等的贡献可以表示为:

$$s_k^*(I) = \frac{\text{cov}(Y^k, Y)}{\sigma^2(Y)}$$

其中 I 是任何一种不平等指数,它当然可以是基尼系数, Y^k 是单项 k 的收入, Y 是总收入,即 $Y = \sum_k Y^k$, $\sigma^2(Y)$ 是 Y 的方差, $\text{cov}(Y^k, Y)$ 是 Y^k 与 Y 的协方差。Shorrocks 的这一理论结果的应用似乎不太广泛,原因是需要 Y^k 与 Y 的二元统计分布,而估计这一分布时需要数据支持,因此其应用范围受到限制。当然,样本数据可用的条件下不存在这种困难,可以用样本方差与样本协方差来进行估计。表 4 的最后一行列出了按上式计算的 4 个单项对总收入不平等的贡献,都是通过样本直接计算的。可见贡献最大的是工资性收入,其次是转移性收入,再次是经营性收入,财产性收入的贡献排在最后。

表 4 2012 年湖北省城镇收入及分解

P_i	总收入		工资性收入		经营性收入		财产性收入		转移性收入	
	L_i	\hat{L}_i	L_i	\hat{L}_i	L_i	\hat{L}_i	L_i	\hat{L}_i	L_i	\hat{L}_i
0.10	0.041	0.041	0.022	0.023	0.000	0.000	0.000	0.000	0.003	0.005
0.20	0.097	0.097	0.065	0.065	0.000	0.000	0.000	0.000	0.020	0.022
0.30	0.162	0.162	0.122	0.122	0.000	0.000	0.000	0.000	0.053	0.054
0.40	0.237	0.237	0.192	0.191	0.000	0.003	0.000	0.000	0.102	0.101
0.50	0.321	0.321	0.274	0.274	0.000	0.011	0.000	0.001	0.168	0.167
0.60	0.416	0.417	0.369	0.370	0.017	0.036	0.003	0.007	0.252	0.252
0.70	0.525	0.525	0.482	0.482	0.091	0.094	0.027	0.029	0.362	0.362
0.80	0.650	0.649	0.614	0.614	0.230	0.219	0.099	0.096	0.501	0.502
0.90	0.795	0.795	0.771	0.771	0.464	0.462	0.279	0.280	0.690	0.689
0.95	0.880	0.880	0.865	0.865	0.647	0.651	0.464	0.464	0.813	0.813
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
平均收入(元)	22 229		13 912		2 130		497		5 690	
MSE $\times 10^6$		0.059		0.308		64.582		3.778		1.795
MAS		0.001		0.001		0.019		0.005		0.003
基尼系数 (估计)		0.256		0.324		0.754		0.853		0.478
基尼系数 (样本)		0.256		0.324		0.758		0.850		0.478
不平等贡献		1.000		0.611		0.113		0.039		0.236

表 5 列出了 2012 年湖北省农村收入分配的估计,其中各项收入都是纯收入。可见除对工资性收入的估计误差稍大外,其他各项的误差可以忽略不计,说明模型 $L_{II}(p)$ 的精度非常理想。从平均收入行可见农村平均收入比城镇低很多(尽管城乡收入概念不同)。农村总纯收入的基尼系数达到了 0.346,大于城镇总收入的基尼系数,因此,可以判断农村的收入不平等大于城镇。农村各收入单项对总不平等的贡献与城镇不同,贡献最大的是经营性收入,其次是工资性收入,财产性收入与转移性收入的贡献都很小。有 20% 的农村人口没有工资性收入,这是其基尼系数很大的原因之一。农村财产性收入的基尼系数达到 0.95,观察洛伦兹曲线的估计,可见是由于只有 20% 的农村人口拥有财产性收入。财产性收入的占比也很小,平均只

有86元。与表4比较,城镇人口中拥有财产性收入的人口比例要大于农村。与城镇比较,值得注意的是农村人口的转移性收入还很少,平均不到500元,这意味着诸如社会保障对农村人口的收入的影响也很小。农村工资性收入与经营性收入的基尼系数都很大,很可能与农村人口的就业分布有关,例如有些人外出打工,从而只有工资性收入;有些人专门从事农业经营,从而只有经营性收入。

表5 2012年湖北省农村收入及分解

P_i	总收入		工资性收入		经营性收入		财产性收入		转移性收入	
	L_i	\hat{L}_i	L_i	\hat{L}_i	L_i	\hat{L}_i	L_i	\hat{L}_i	L_i	\hat{L}_i
0.10	0.024	0.024	0.000	0.001	0.007	0.007	0.000	0.000	0.000	0.001
0.20	0.065	0.065	0.004	0.009	0.024	0.024	0.000	0.000	0.004	0.006
0.30	0.119	0.119	0.024	0.029	0.052	0.052	0.000	0.000	0.014	0.015
0.40	0.184	0.184	0.062	0.064	0.094	0.093	0.000	0.000	0.030	0.031
0.50	0.262	0.262	0.122	0.120	0.150	0.150	0.000	0.000	0.055	0.055
0.60	0.353	0.353	0.204	0.201	0.225	0.226	0.000	0.000	0.090	0.090
0.70	0.460	0.460	0.315	0.313	0.326	0.326	0.000	0.000	0.145	0.144
0.80	0.586	0.587	0.460	0.462	0.460	0.459	0.005	0.005	0.228	0.227
0.90	0.747	0.747	0.658	0.661	0.643	0.643	0.052	0.052	0.376	0.378
0.95	0.849	0.849	0.793	0.791	0.773	0.773	0.161	0.161	0.530	0.529
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
平均收入(元)	8 890		3 612		4 825		86		499	
MSE × 10 ⁶		0.036		9.447		0.162		0.016		1.631
MAS		0.000		0.003		0.000		0.000		0.001
基尼系数(估计)		0.346		0.537		0.515		0.951		0.738
基尼系数(样本)		0.346		0.539		0.515		0.950		0.740
不平等贡献		1.000		0.250		0.706		0.018		0.041

图1(A)与图1(B)描述了城乡洛伦兹曲线估计,其中曲线上的“+”表示原始数据点,可见估计的洛伦兹曲线基本都准确通过这些点。两图直观反映了我国城乡人口的收入不平等的结构不同。

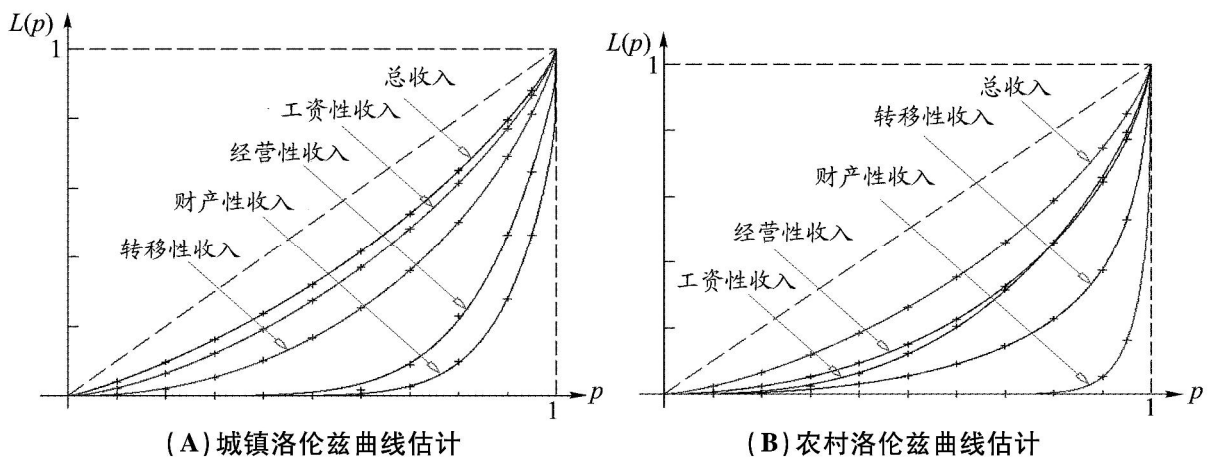


图1

考虑到文章篇幅,不对2009年与2010年列出类似于表4与表5的洛伦兹曲线估计,表6列出了这两年城乡基尼系数估计与各单项收入对总收入不平等的贡献。表7列出了表4与表5中各条洛伦兹曲线的参数估计。2009年与2010年与表4、表5类似的数据可向本文作者索取。

表6 2009-2010年湖北省城乡收入不平等估计

	总收入	工资性收入	经营性收入	财产性收入	转移性收入
2009年城镇					
基尼系数(估计)	0.252	0.327	0.826	0.878	0.517
基尼系数(样本)	0.252	0.328	0.826	0.876	0.518
不平等贡献	1.000	0.565	0.217	0.023	0.194
2010年城镇					
基尼系数(估计)	0.271	0.372	0.848	0.912	0.550
基尼系数(样本)	0.271	0.371	0.846	0.910	0.550
不平等贡献	1.000	0.652	0.138	0.039	0.170
2009年农村					
基尼系数(估计)	0.334	0.538	0.461	0.888	0.676
基尼系数(样本)	0.334	0.540	0.461	0.888	0.677
不平等贡献	1.000	0.280	0.673	0.014	0.044
2010年农村					
基尼系数(估计)	0.341	0.529	0.472	0.918	0.698
基尼系数(样本)	0.341	0.531	0.472	0.921	0.698
不平等贡献	1.000	0.273	0.615	0.081	0.037

表7 2012年洛伦兹曲线的参数估计

	ω	τ	δ	γ	α	η
城镇						
总收入	0.5231	0.6035	0.1864	0.9984	0.0528	1.1290
工资性收入	0.9975	0.5841	0.0000	0.9986	0.0613	1.4497
经营性收入	0.0000	1.0000	0.9529	0.0000	0.9029	2.7098
财产性收入	1.0000	0.0002	0.1815	0.0000	6.7884	2.1364
转移性收入	1.0000	0.9458	0.0589	0.0000	1.6467	0.5000
农村						
总收入	0.9265	0.9990	0.9001	0.0000	0.5144	0.5222
工资性收入	0.9909	0.0000	0.0954	0.9920	0.0001	2.7415
经营性收入	0.5460	0.9869	0.4796	0.5468	0.7116	0.8313
财产性收入	0.0004	0.9993	0.8291	0.2239	0.3791	9.4746
转移性收入	0.9789	0.9894	0.0000	0.0000	0.9231	1.4384

由表6可见城乡各种收入分量的不平等状态与2012年相差不大,不同收入分量对整个收入不平等的贡献格局也基本相同。

注意到,从上述估计结果看,2012年农村总收入的基尼系数大约为0.35左右,而城镇总收入的基尼系数只有0.26左右,因此城镇总收入的不平等看来并不严重。但这里需要强调两点,一是城镇部分的数据采集限定的是户籍人口,没有包括已经进城的农村流动人口;二是我国城镇数据调查中,高收入人口参与意愿一般不高,调查户中可能一般是中低收入人口,数据描述的收入分配的代表性可能存在疑问。因此,我国收入分配数据的准确性可能需要提高。

六、结语

本文先提出了三种单参数洛伦兹曲线模型,再在此基础上构造了其他四种模型,用这些构造得到的模型对其他国家或地区的数据进行的数据实验,说明所提出的模型效果良好。文中再用效果可能最好的模型对湖北省城乡收入分配数据进行了测算。由于湖北省的数据中包含了各个主要的收入分量,可以考察各种收入分量对总收入不平等的贡献。这一分析说明湖北省城乡财产性收入占比还非常小,只有少量人口拥有财产性收入,因此财产性收入的不平等程度极大,但对总收入不平等的贡献还不小。对城镇收入不平等贡献最大的收入分量是工资性收入与转移性收入,对农村收入不平等贡献最大的是经营性收入与工资性收入,农村人口的转移性收入占比对总收入不平等的贡献还很小。因此,调节我国城乡居民收入分配的政策可以如下:其一,增加城乡普通民众的财产性收入,注意调节财产性收入的公平分配。其二,增加农村转移性收入,例如增加农村社会保障制度的覆盖面、增加保障力度。其三,农村居民收入的提高主要靠工资性收入与经营性收入,因此在促进农村剩余劳动力转移的同时,应千方百计增加农村从业人员的收入。其四,城镇人口的经营性收入有提高的空间,这可以通过促进第三产业的发展来实现。

本文理论模型 $L_{II}(p)$ 可能非常适用于我国的收入分配数据,进一步研究时可以考察这一模型对其他领域数据的适用性。我国很多部门都出版分组数据,人口普查、经济普查等数据中也存在大量分组数据,这为本模型的使用提供了很大空间。还可能考虑使用本文方法研究其他洛伦兹曲线模型。

参考文献:

1. 张奎、王原君,2010:《Sarabia 洛伦兹曲线模型的推广》,《应用数学》第3期。
2. Basmann, R. L., K. J. Hayes, and D. J. Slottje. 1990. "A General Functional Form for Approximating the Lorenz Curve." *Journal of Econometrics*, 43(1-2): 77-90.
3. Basmann, R. L., K. J. Hayes, and D. J. Slottje. 1991. "The Lorenz Curve and the Mobility Function." *Economics Letters*, 36(1): 105-111.
4. Chotikapanich, D. 1993. "A Comparison of Alternative Functional Forms for the Lorenz Curve." *Economics Letters*, 4(12): 187-192.
5. Chotikapanich, D., B. Griffiths, and D. S. Rao. 2007. "Estimating and Combining National Income Distributions Using Limited Data." *Journal of Business and Economic Statistics*, 25(1): 97-109.
6. Gupta, M. R. 1984. "Functional Form for Estimating the Lorenz Curve." *Econometrica*, 52(5): 1313-1314.
7. Kakwani, N. C. 1980. "Functional Forms for Estimating the Lorenz Curve; A Reply." *Econometrica*, 48(4): 1063-1064.
8. Kakwani, N. C., and N. Podder. 1973. "Efficient Estimation of Lorenz Curve and Associated Inequality Measures from Grouped Observations." *Econometrica*, 44(1): 137-148.
9. Ogwang, T., and U. L. Gouranga Rao. 1996. "A New Functional Form for Approximating the Lorenz Curve." *Economics Letters*, 52(1): 21-29.
10. Ogwang, T., and U. L. Gouranga Rao. 2000. "Hybrid Models of the Lorenz Curve." *Economics Letters*, 69(1): 39-44.
11. Ortega, P., G. Martin, A. Fernandez, M. Ladoux, and A. Garcia. 1991. "A New Functional Form for Estimating Lorenz Curves." *Review of Income and Wealth*, 37(4): 447-452.
12. Rasche, R. H., J. Gaffney, and A. Y. C. Koo. 1980. "Functional Forms for Estimating the Lorenz Curve." *Econometrica*, 48(4): 1061-1062.
13. Rasche, R. H., J. Gaffney, A. Y. C. Koo, and N. Obst. 1984. "Functional Form for Estimating the Lorenz Curve." *Econometrica*, 52(5): 1341-1342.
14. Rohde, N. 2009. "An Alternative Functional Form for Estimating the Lorenz Curve." *Economics Letters*, 105(1): 61-63.

(下转第27页)

state owned proportion significantly reduces urban concentration rate, and the robust tests further support our hypothesis. The finding means that state control relaxing is a feasible way of effective urban system construction.

Key Words: Urban System; State – owned Economy; City Size Distribution; Urbanization

JEL Classification: R12

(责任编辑:赵锐)

(上接第15页)

15. Ryn, H. K., and D. J. Slottje. 1996. "Two Flexible Functional Form Approaches for Approximating the Lorenz Curve." *Journal of Econometrics*, 72(1-2): 251-274.
16. Sarabia, J. 1997. "A Hierarchy of Lorenz Curves Based on the Generalized Tukey's Lambda Distribution." *Econometric Reviews*, 16(2): 305-320.
17. Sarabia, J., E. Castillo, and D. J. Slottje. 1999. "An Ordered Family of Lorenz Curves." *Journal of Econometrics*, 91(1): 43-60.
18. Sarabia, J., E. Castillo, and D. J. Slottje. 2001. "An Exponential Family of Lorenz Curves." *Southern Economic Journal*, 67(3): 748-756.
19. Sarabia, J., E. Castillo, M. Pascual, and M. Sarabia. 2005. "Mixture Lorenz Curves." *Economics Letters*, 89(1): 89-94.
20. Schader, M., and F. Schmid. 1994. "Fitting Parametric Lorenz Curves to Grouped Income Distribution: A Critical Note." *Empirical Economics*, 19: 361-370.
21. Shorrocks, A. F. 1982. "Inequality Decomposition by Factor Components." *Econometrica*, 50(1): 193-212.
22. Villasenor, J. A., and B. C. Arnold. 1989. "Elliptical Lorenz Curves." *Journal of Econometrics*, 40(2): 327-338.
23. Wang, Z. X., R. Smyth, and Ng Y-K. 2011. "A General Method for Creating Lorenz Curves." *Review of Income and Wealth*, 57(3): 561-582.
24. Wang, Zuxiang, and R. Smyth. 2013. "A Hybrid Method for Creating Lorenz Curves with an Application to Measuring World Income Inequality." Monash Economics Department Discussion Paper, No. 46/13.

A New Lorenz Model for Income Distribution Analysis: Model Construction and Application

You Shibing and Wang Yuanjun

(Economics and Management School of Wuhan University)

Abstract: The research of Lorenz Curve model is an active research field in the income distribution analysis. This paper puts forward several effective models, and applies these models to international data widely used in the literature, showing that the new models are far better than those models which are used in the literature. Then the best model found is used to study the income distribution of China. This paper uses 2009, 2010, 2012 income distribution sample data of Hubei province, forming grouped data, then uses one of the new Lorenz Curve models to fit the grouped data, estimates the Gini coefficient, and proves that the resulting Gini coefficient is very close to the Gini coefficient calculated from the sample data. The estimated Lorenz Curve almost accurately passes the grouped data points. So the proposed Lorenz Curve model has important application value, which is suitable for the study of income distribution of our country. The analysis of income distribution of the province shows that the property income inequality is the most serious, and some population have no property income at all. The number or proportion of rural transfer income still lags behind that of the urban area.

Key Words: Lorenz Curve; Income Distribution; Gini Coefficient

JEL Classification: A11, C63, J31

(责任编辑:彭爽)